Microbiological Reviews

A Publication of the American Society for Microbiology

VOLUME 57 • SEPTEMBER 1993 • NUMBER 3

CONTENTS/SUMMARIES

Summary: The provirus structure of retroviruses is bracketed by long terminal repeats (LTRs). The two LTRs (5' and 3') are identical in nucleotide sequence and organization. They contain signals for transcription initiation as well as termination and cleavage/polyadenylation. As in eukaryotic pre-mRNAs, the two common signals, the polyadenylation signal, AAUAAA, or a variant AGUAAA, and the G+U-rich sequence are present in all retroviruses. However, the AAUAAA sequence is present in the U3 region in some retroviruses and in the R region in other retroviruses. As in animal cell RNAs, both AAUAAA and G+U-rich sequences apparently contribute to the 3'-end processing of retroviral RNAs. In addition, at least in a few cases examined, the sequences in the U3 region determine the efficiency of 3'-end processing. In retroviruses in which the AAUAAA is localized in the R region, the poly(A) signal in the 3' LTR but not the 5' LTR must be selectively used for the production of genomic RNA. It appears that the short distance between the 5' cap site and polyadenylation signal in the 5' LTR precludes premature termination and polyadenylation. Since 5' and 3' LTRs are identical in sequence and structural organization yet function differently, it is speculated that flanking cellular DNA sequences, chromatin structure, and binding of transcription factors may be involved in the functional divergence of 5' and 3' LTRs of retroviruses.

Summary: Our understanding of fatty acid biosynthesis in Escherichia coli has increased greatly in recent years. Since the discovery that the intermediates of fatty acid biosynthesis are bound to the heat-stable protein cofactor termed acyl carrier protein, the fatty acid synthesis pathway of E. coli has been studied in some detail. Interestingly, many advances in the field have aided in the discovery of analogous systems in other organisms. In fact, E. coli has provided a paradigm of predictive value for the synthesis of fatty acids in bacteria and plants and the synthesis of bacterial polyketide antibiotics. In this review, we concentrate on four major areas of research. First, the reactions in fatty acid biosynthesis and the proteins catalyzing these reactions are discussed in detail. The genes encoding many of these proteins have been cloned, and characterization of these genes has led to a better understanding of the pathway. Second, the function and role of the two essential cofactors in fatty acid synthesis, coenzyme A and

Continued on following page

Continued from preceding page

acyl carrier protein, are addressed. Finally, the steps governing the spectrum of products produced in synthesis and alternative destinations, other than membrane phospholipid, for fatty acids in E. coli are described. Throughout the review, the contribution of each portion of the pathway to the global regulation of synthesis is examined. In no other organism is the bulk of knowledge regarding fatty acid metabolism so great; however, questions still remain to be answered. Pursuing such questions should reveal additional regulatory mechanisms of fatty acid synthesis and, hopefully, the role of fatty acid synthesis and other cellular processes in the global control of cellular growth.

Summary: Numerous gram-negative and gram-positive bacteria take up carbohydrates through the phosphoenolpyruvate (PEP):carbohydrate phosphotransferase system (PTS). This system transports and phosphorylates carbohydrates at the expense of PEP and is the subject of this review. The PTS consists of two general proteins, enzyme I and HPr, and a number of carbohydrate-specific enzymes, the enzymes II. PTS proteins are phosphoproteins in which the phospho group is attached to either a histidine residue or, in a number of cases, a cysteine residue. After phosphorylation of enzyme I by PEP, the phospho group is transferred to HPr. The enzymes II are required for the transport of the carbohydrates across the membrane and the transfer of the phospho group from phospho-HPr to the carbohydrates. Biochemical, structural, and molecular genetic studies have shown that the various enzymes II have the same basic structure. Each enzyme II consists of domains for specific functions, e.g., binding of the carbohydrate or phosphorylation. Each enzyme II complex can consist of one to four different polypeptides. The enzymes II can be placed into at least four classes on the basis of sequence similarity. The genetics of the PTS is complex, and the expression of PTS proteins is intricately regulated because of the central roles of these proteins in nutrient acquisition. In addition to classical induction-repression mechanisms involving repressor and activator proteins, other types of regulation, such as antitermination, have been observed in some PTSs. Apart from their role in carbohydrate transport, PTS proteins are involved in chemotaxis toward PTS carbohydrates. Furthermore, the IIA^{Glc} protein, part of the glucose-specific PTS, is a central regulatory protein which in its nonphosphorylated form can bind to and inhibit several non-PTS uptake systems and thus prevent entry of inducers. In its phosphorylated form, $P\text{-}IIA^{Glc}$ is involved in the activation of adenylate cyclase and thus in the regulation of gene expression. By sensing the presence of PTS carbohydrates in the medium and adjusting the phosphorylation state of IIA^{Glc}, cells can adapt quickly to changing conditions in the environment. In gram-positive bacteria, it has been demonstrated that HPr can be phosphorylated by ATP on a serine residue and this modification may perform a regulatory function.

Summary: Several species of the genus Fusarium and related fungi produce trichothecenes which are sesquiterpenoid epoxides that act as potent inhibitors of eukaryotic protein synthesis. Interest in the trichothecenes is due primarily to their widespread contamination of agricultural commodities and their adverse effects on human and animal health. In this review, we describe the trichothecene biosynthetic pathway in Fusarium species and discuss genetic evidence that several trichothecene biosynthetic genes are organized in a gene cluster. Trichothecenes are highly toxic to a wide range of eukaryotes, but their specific function, if any, in the survival of the fungi that produce them is not obvious. Trichothecene gene disruption experiments indicate that production of trichothecenes can enhance the severity of disease caused by Fusarium species on some plant hosts. Understanding the regulation and function of trichothecene biosynthesis may aid in development of new strategies for controlling their production in food and feed products.

Continued on following page

Continued from preceding page

Summary: The white rot basidiomycete Phanerochaete chrysosporium completely degrades lignin and a variety of aromatic pollutants during the secondary metabolic phase of growth. Two families of secreted heme enzymes, lignin peroxidase (LiP) and manganese peroxidase (MnP), are major components of the extracellular lignin degradative system of this organism. MnP and LiP both are encoded by families of genes, and the lip genes appear to be clustered. The lip genes contain eight or nine short introns; the mnp genes contain six or seven short introns. The sequences surrounding active-site residues are conserved among LiP, MnP, cytochrome c peroxidase, and plant peroxidases. The eight LiP cysteine residues align with 8 of the 10 cysteines in MnP. LiPs are synthesized as preproenzymes with a 21-amino-acid signal sequence followed by a 6- or 7-amino-acid propeptide. MnPs have a 21- or 24-amino-acid signal sequence but apparently lack a propeptide. Both LiP and MnP are regulated at the mRNA level by nitrogen, and the various isozymes may be differentially regulated by carbon and nitrogen. MnP also is regulated at the level of gene transcription by Mn(II), the substrate for the enzyme, and by heat shock. The promoter regions of mnp genes contain multiple heat shock elements as well as sequences that are identical to the consensus metal regulatory elements found in mammalian metallothionein genes. DNA transformation systems have been developed for P. chrysosporium and are being used for studies on gene regulation and for gene replacement experiments.

Summary: Several data libraries have been created to organize all the data obtained worldwide about the Escherichia coli genome. Because the known data now amount to more than 40% of the whole genome sequence, it has become necessary to organize the data in such a way that appropriate procedures can associate knowledge produced by experiments about each gene to its position on the chromosome and its relation to other relevant genes, for example. In addition, global properties of genes, affected by the introduction of new entries, should be present as appropriate description fields. A data base, implemented on Macintosh by using the data base management system 4th Dimension, is described. It is constructed around a core constituted by known contigs of E. coli sequences and links data collected in general libraries (unmodified) to data associated with evolving knowledge (with modifiable fields). Biologically significant results obtained through the coupling of appropriate procedures (learning or statistical data analysis) are presented. The data base is available through a 4th Dimension runtime and through FTP on Internet. It has been regularly updated and will be systematically linked to other E. coli data bases (M. Kroger, R. Wahl, G. Schachtel, and P. Rice, Nucleic Acids Res. 20(Suppl.):2119–2144, 1992; K. E. Rudd, W. Miller, C. Werner, J. Ostell, C. Tolstoshev, and S. G. Satterfield, Nucleic Acids Res. 19:637–647, 1991) in the near future.

Summary: From a historical perspective, the study of both the biochemistry and the genetics of lipopolysaccharide (LPS) synthesis began with the enteric bacteria. These organisms have again come to the forefront as the blocks of genes involved in LPS synthesis have been sequenced and analyzed. A number of new and unanticipated genes were found in these clusters, indicating a complexity of the biochemical pathways which was not predicted from the older studies. One of the most dramatic areas of LPS research has been the elucidation of the lipid A biosynthetic pathway. Four of the genes in this pathway have now been identified and sequenced, and three of them are located in a complex operon which also contains genes involved in DNA and phospholipid

Continued on following page

Continued from preceding page

synthesis. The rfa gene cluster, which contains many of the genes for LPS core synthesis, includes at least 17 genes. One of the remarkable findings in this cluster is a group of several genes which appear to be involved in the synthesis of alternate rough core species which are modified so that they cannot be acceptors for O-specific polysaccharides. The rfb gene clusters which encode O-antigen synthesis have been sequenced from a number of serotypes and exhibit the genetic polymorphism anticipated on the basis of the chemical complexity of the O antigens. These clusters appear to have originated by the exchange of blocks of genes among ancestral organisms. Among the large number of LPS genes which have now been sequenced from these rfa and rfb clusters, there are none which encode proteins that appear to be secreted across the cytoplasmic membrane and surprisingly few which encode integral membrane proteins or proteins with extensive hydrophobic domains. These data, together with sequence comparison and complementation experiments across strain and species lines, suggest that the LPS biosynthetic enzymes may be organized into clusters on the inner surface of the cytoplasmic membrane which are organized around a few key membrane proteins.

Summary: Temperate coliphage P2 and satellite phage P4 have icosahedral capsids and contractile tails with side tail fibers. Because P4 requires all the capsid, tail, and lysis genes (late genes) of P2, the genomes of these phages are in constant communication during P4 development. The P4 genome (11,624 bp) and the P2 genome (33.8 kb) share homologous cos sites of 55 bp which are essential for generating 19-bp cohesive ends but are otherwise dissimilar. P4 turns on the expression of helper phage late genes by two mechanisms: derepression of P2 prophage and transactivation of P2 late-gene promoters. P4 also exploits the morphopoietic pathway of P2 by controlling the capsid size to fit its smaller genome. The P4 sid gene product is responsible for capsid size determination, and the P2 capsid gene product, gpN, is used to build both sizes. The P2 capsid contains 420 capsid protein subunits, and P4 contains 240 subunits. The size reduction appears to involve a major change of the whole hexamer complex. The P4 particles are less stable to heat inactivation, unless their capsids are coated with a P4-encoded decoration protein (the psu gene product). P4 uses a small RNA molecule as its immunity factor. Expression of P4 replication functions is prevented by premature transcription termination effected by this small RNA molecule, which contains a sequence that is complementary to a sequence in the transcript that it terminates.

Summary: The transcription of nucleus-encoded genes in eukaryotes is performed by three distinct RNA polymerases termed I, II, and III, each of which is a complex enzyme composed of more than 10 subunits. The isolation of genes encoding subunits of eukaryotic RNA polymerases from a wide spectrum of organisms has confirmed previous biochemical and immunological data indicating that all three enzymes are closely related in structures that have been conserved in evolution. Each RNA polymerase is an enzyme complex composed of two large subunits that are homologous to the two largest subunits of prokaryotic RNA polymerases and are associated with smaller polypeptides, some of which are common to two or to all three eukaryotic enzymes. This remarkable conservation of structure most probably underlies a conservation of function and emphasizes the likelihood that information gained from the study of RNA polymerases from one organism will be applicable to others. The recent isolation of many mutations affecting the structure and/or function of eukaryotic and prokaryotic RNA polymerases now makes it feasible to begin integrating genetic and biochemical information from various species in order to develop a picture of these enzymes. The picture of eukaryotic RNA polymerases depicted in this article emphasizes the role(s) of different polypeptide regions in interaction with other subunits, cofactors, substrates, inhibitors, or accessory transcription factors, as well as the

Continued on following page

Continued from preceding page

requirement for these interactions in transcription initiation, elongation, pausing, termination, and/or enzyme assembly. Most mutations described here have been isolated in eukaryotic organisms that have well-developed experimental genetic systems as well as amenable biochemistry, such as Saccharomyces cerevisiae, Drosophila melanogaster, and Caenorhabditis elegans. When relevant, mutations affecting regions of Escherichia coli RNA polymerase that are conserved among eukaryotes and prokaryotes are also presented. In addition to providing information about the structure and function of eukaryotic RNA polymerases, the study of mutations and of the pleiotropic phenotypes they imposed has underscored the central role played by these enzymes in many fundamental processes such as development and cellular differentiation.

Summary: Photosynthetic organisms can acclimate to their environment by changing many cellular processes, including the biosynthesis of the photosynthetic apparatus. In this article we discuss the phycobilisome, the light-harvesting apparatus of cyanobacteria and red algae. Unlike most light-harvesting antenna complexes, the phycobilisome is not an integral membrane complex but is attached to the surface of the photosynthetic membranes. It is composed of both the pigmented phycobiliproteins and the nonpigmented linker polypeptides; the former are important for absorbing light energy, while the latter are important for stability and assembly of the complex. The composition of the phycobilisome is very sensitive to a number of different environmental factors. Some of the filamentous cyanobacteria can alter the composition of the phycobilisome in response to the prevalent wavelengths of light in the environment. This process, called complementary chromatic adaptation, allows these organisms to efficiently utilize available light energy to drive photosynthetic electron transport and CO_2 fixation. Under conditions of macronutrient limitation, many cyanobacteria degrade their phycobilisomes in a rapid and orderly fashion. Since the phycobilisome is an abundant component of the cell, its degradation may provide a substantial amount of nitrogen to nitrogen-limited cells. Furthermore, degradation of the phycobilisome during nutrient-limited growth may prevent photodamage that would occur if the cells were to absorb light under conditions of metabolic arrest. The interplay of various environmental parameters in determining the number of phycobilisomes and their structural characteristics and the ways in which these parameters control phycobilisome biosynthesis are fertile areas for investigation.

Summary: Treponema pallidum subsp. pallidum, the spirochete that causes syphilis, is unusual in a number of respects, including its small genome size, inability to grow under standard in vitro culture conditions, microaerophilism, apparent paucity of outer membrane proteins, structurally complex periplasmic flagella, and ability to evade the host immune responses and cause disease over a period of years to decades. Many of these attributes are related ultimately to its protein content. Our knowledge of the activities, structure, and immunogenicity of its proteins has been expanded by the application of recombinant DNA, hybridoma, and structural fractionation techniques. The purpose of this monograph is to summarize and correlate this new information by using two-dimensional gel electrophoresis, monoclonal antibody reactivity, sequence data, and other properties as the bases of polypeptide identification. The protein profiles of the T. pallidum subspecies causing syphilis, yaws, and endemic syphilis are virtually indistinguishable but differ considerably from those of other treponemal species. Among the most abundant polypeptides are a group of lipoproteins of unknown function that appear to be important in the immune response during syphilitic infection.

Continued on following page

Continued from preceding page

The periplasmic flagella of T. pallidum and other spirochetes are unique with regard to their protein content and ultrastructure, as well as their periplasmic location. They are composed of three core proteins (homologous to the other members of the eubacterial flagellin family) and a single, unrelated sheath protein; the functional significance of this arrangement is not understood at present. Although the bacterium contains the chaperonins GroEL and DnaK, these proteins are not under the control of the heat shock regulon as they are in most organisms. Studies of the immunogenicity of T. pallidum proteins indicate that many may be useful for immunodiagnosis and immunoprotection. Future goals in T. pallidum polypeptide research include continued elucidation of their structural locations and functional activities, identification and characterization of the low-abundance outer membrane proteins, further study of the immunoprotective and immunodiagnostic potential of T. pallidum proteins, and clarification of the roles of treponemal proteins in pathogenesis.